VARIABLE-FREQUENCY SYNTHESIS: AN IMPROVED HARMONIC CODING SCHEME

Luís B. Almeida and Fernando M. Silva

Centro de Análise e Processamento de Sinais

Instituto Superior Técnico

1096 LISBOA CODEX - PORTUGAL

ABSTRACT

The Harmonic Coding concept has already shown its potential for efficiently coding speech. Previous implementations have used a frame rate of one every 16 ms. This was mainly due to the fact that, with longer frames, even a nonstationary spectral model (of low order) cannot reproduce the zones of fast-varying pitch with the desirable quality. However, the high framing rate is a limitation, since it implies that fewer bits will be available for encoding each frame.

A solution for this problem has been devised: the signal is synthesized in the time domain, as a superimposition of "harmonics" whose instantaneous frequency varies continuously along an interpolation curve, within each frame. In this way, fast pitch variations can be tracked with no difficulty. Experimental results are presented, confirming these facts. The integration of this synthesis scheme in a speech coder is discussed.

INTRODUCTION

Harmonic Coding (HC) is a very efficient and very flexible speech coding technique. The original HC scheme was described in [1,2,3]. The basic analysis-synthesis scheme underlying HC relies on a spectral model for voiced speech [2,3,4]. In the implementations done so far, this has simply been the classical spectral line model. In these implementations, a 32 ms-long Hanning window was used both in the analysis and in the synthesis, with a 50% overlap of consecutive frames, resulting in a frame rate of one every 16 ms.

From the experience gained with those implementations, it was concluded that a lower frame rate (about one every 32 ms) was desirable, in order to allow more bits for the coding of each frame. However, when this lower frame rate was tried, using just the basic analysis-synthesis (with no quantization), it was found that segments with fast-varying pitch were not well reproduced: the pitch appeared to jump in very fast (but discernible) steps, instead of evolving smoothly. Figure 1-a) and b) show an example of this situation: the steps in the frequencies of the harmonics are clearly visible in the left part of figure 1-b).

The reason for this phenomenon is quite simple: the spectral line model corresponds, in the time domain, to a purely periodic waveform. Thus, the pitch of the output signal is constant, within each synthesis frame; the overlap between consecutive frames does not yield a smooth pitch variation, as the figure shows. The nonstationary model described in [2,3,4] yields good reproduction of pitch variations if one allows a high model order. However, for efficient coding, the order has to be bounded to a low value (say, 2). Analysis-synthesis based on a 2nd-order model showed that the steppy character of the pitch evolution was somewhat reduced, but still perceptible.

THE NEW SYNTHESIS SCHEME

The Variable-Frequency Synthesis (VFS) scheme, first introduced in [5], was devised as a solution to this problem. In this scheme, each synthesis frame consists of the segment between the centers of two consecutive analysis frames (figure 2-a). The harmonic analysis gives us the amplitude and phase of each harmonic at both ends of this segment. The synthesis is performed by adding, in the time domain, "harmonics" of continuously varying amplitude and phase. The amplitude evolution along the segment is simply obtained by linear interpolation between the values found at both ends (figure 2-b). The phase evolution is given by a 3rd-degree polynomial (with time as free variable), as shown in figure 3-c). The coefficients of this polynomial are such that, at each end of the segment, the phase equals the value found in the analysis, and its time derivative equals the instantaneous frequency of the harmonic, which is simply the fundamental frequency at that point multiplied by the order of the harmonic (in fact, this is only an approximate value of the instantaneous frequency - cf. [3]). The instantaneous frequency of the synthesized harmonic, being the time derivative of the phase, is given by a 2nd- degree polynomial in time, and can thus easily follow fast pitch variations. This synthesis method guarantees the continuity of both the signal and its derivative across synthesis frame boundaries.

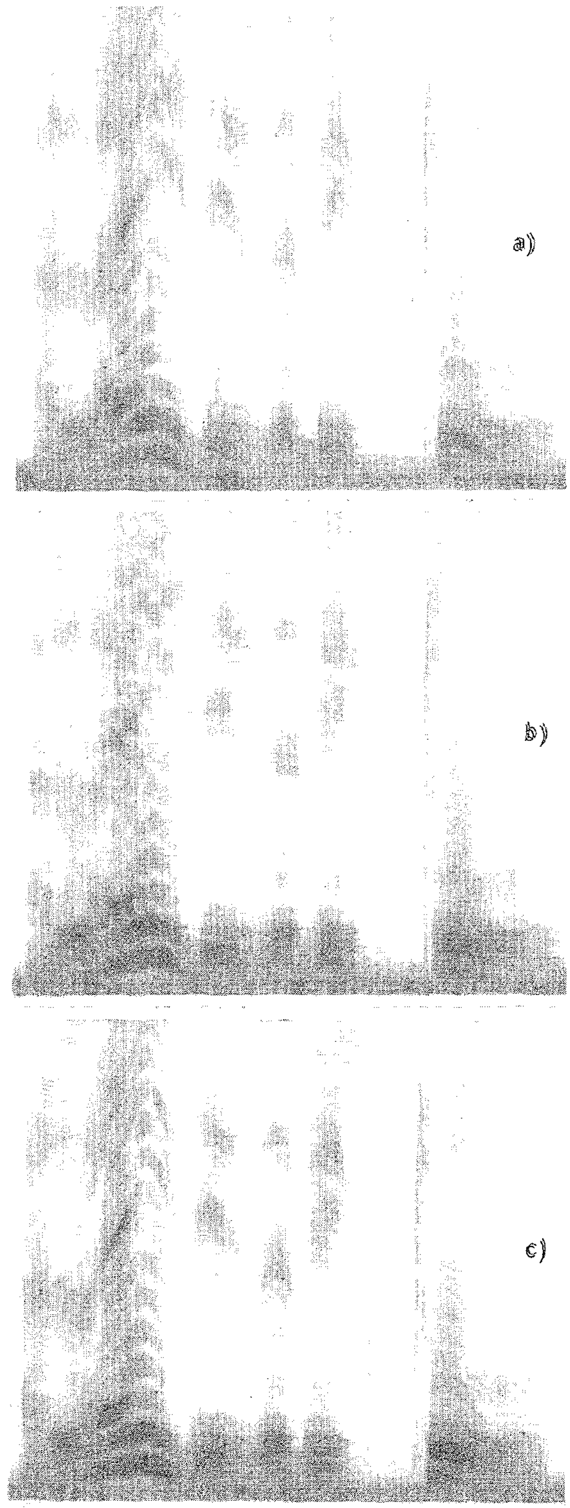


Figure 1 - Spectrograms showing the performance of analysis-synthesis schemes.
a) Input signal. b) Output of the previous analysis-synthesis scheme. c) Output of the VFS scheme. The sentence is "a lathe is a big tool", spoken by a male speaker. The frequency range is 0 - 3.5 kHz.

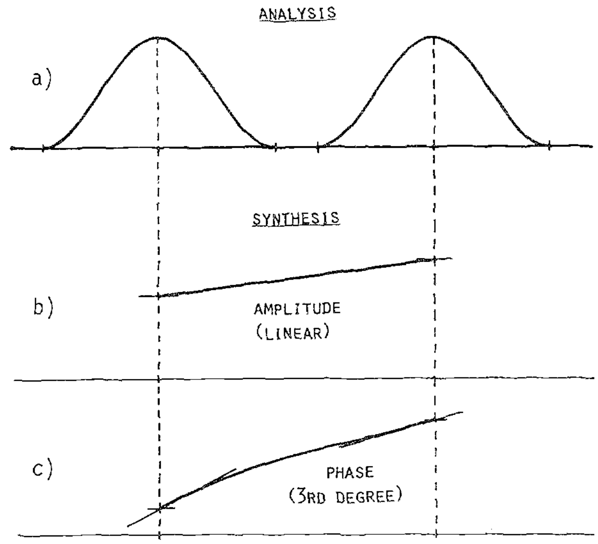


Figure 2 - Illustration of the VFS synthesis method. a) Analysis. b) Amplitude interpolation in the synthesis. c) Phase interpolation in the synthesis.

The main difficulty in this synthesis scheme consists of the fact that the harmonic analysis only supplies the principal (modulo 2π) values of the phases of the harmonics. To find the interpolating polynomial, the actual (as opposed to principal) value of the phase difference between both ends of the synthesis frame must be found. In a simple analysis-synthesis environment, intermediate frames can always be used, if necessary, to find the correct value. However, in a coding application, the transmission of any information concerning the number of 2π intervals means an increase in the bit rate. This problem, which is a kind of phase unwrapping, is closely related to phase prediction [2,3,4], and a few methods for solving it are presently under study.

EXPERIMENTAL RESULTS

Figure 1-c) shows a spectrogram of the output obtained with the VFS scheme. The steps in the harmonic frequencies have completely disappeared. This fact was confirmed by informal listening tests which showed that the quality of the VFS output is clearly superior to that of the previous scheme. Table I shows a comparison of segmental SNR for both schemes, averaged over a set of 12 speech records (three sentences, each spoken by four speakers, two male and two female).

Table I

Comparison of segmental SNR's for previous and VFS synthesis schemes. SNR was computed only in voiced segments.

Previous scheme	VFS scheme
8.8 dB	10.5 dB

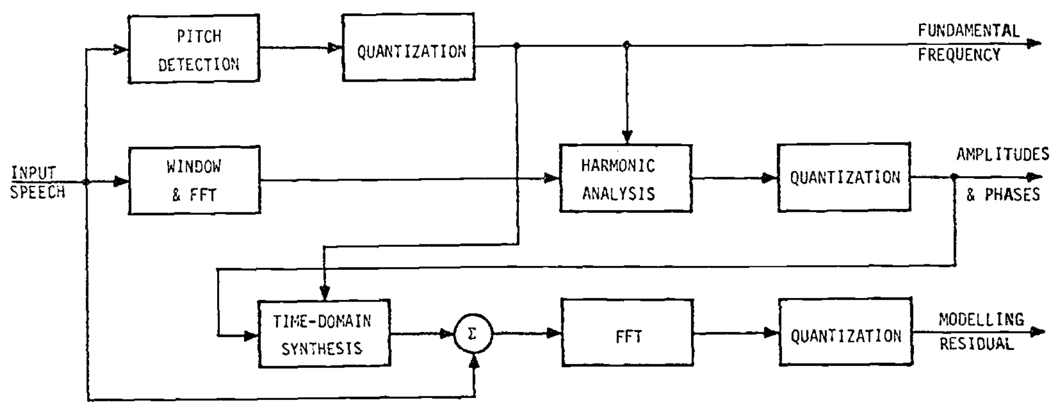
These results also confirm the better synthesis performed by the VFS scheme. It should be noted that, as is well known, segmental SNR is not a good measure of perceptual quality. The perceptual quality of VFS output, as judged by informal listening, is very good, contrarily to what table-I might suggest. The main significance of segmental SNR values, in this context, is as a measure of the relative energy of the modelling error, which is important for coding purposes, as explained in the next section.

PROPOSED CODING SCHEME

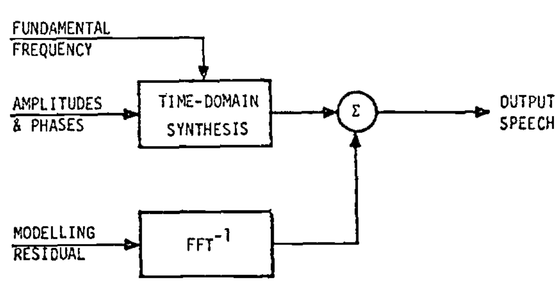
The proposed HC-VFS coder structure is shown in figure 3. The main difference from the previous HC structure (cf. [1,2,3]) lies in the fact that the synthesis is now performed in the time domain, using the VFS scheme. The modelling residual is then found, also in the time domain, by subtracting the synthetic signal from the input speech. This residual is transformed to the frequency domain, and then quantized. At the receiver, the VFS synthesis is repeated, and the modelling residual is added to its output, after being passed to the time domain. Prefiltering at the transmitter input (e.g. dynamic prewhitening, or simple pre-emphasis) can of course be used, the corresponding postfiltering being then incorporated in the receiver.

The HC-VFS scheme uses essentially the same transmitted information as the simple HC scheme: the fundamental frequency and model coefficients (harmonic amplitudes and phases), and the modelling residual. Some further information for the phase unwrapping may however be needed, as mentioned in the previous section. The dynamic bit assignment, discussed in [1,2,3] is also to be used in this coder, in exactly the same fashion. Table I shows that the energy of the modelling residual will be lower than in the previous scheme, suggesting that this residual can be encoded with a lower number of bits.

The amount of processing involved in HC-VFS is somewhat higher than that of simple HC. In fact,



a) Transmitter



b) Receiver

Figure 3 - Generic diagram of the HC-VFS coder

the time-domain synthesis involves more computation than the previous frequency-domain one, and there is one more FFT per frame, in the transmitter.

A side advantage of the new coding scheme is that the lengths of the analysis and synthesis frames are now completely independent from each other: the analysis window can be chosen to best suit the analysis mechanism, and the synthesis frame length can be chosen on the basis of quality and bit rate considerations. This length can even be varied dynamically, for variable bit rate applications.

CONCLUSIONS

A new synthesis scheme for Harmonic Coding was introduced, which is based on the superimposition, in the time domain, of varying-frequency harmonics. Its ability to reproduce segments of fast-varying pitch was experimentally demonstrated. Additionally, this new scheme permits the synthesis frame length to be chosen independently from the analysis window. This affords a further degree of flexibility for the optimization of the coder. The information transmitted in the new coder structure is essentially the same as in the previous one, though it is not known yet whether the transmission of some further information will be needed.

REFERENCES

[1] - L. B. Almeida and J. M. Tribolet, "Harmonic Coding: a Low Bit Rate, Good-Quality Speech Coding Technique", Proc. ICASSP 1982.

[2] - L. B. Almeida and J. M. Tribolet, "Non-Stationary Spectral Modelling of Voiced Speech", IEEE Trans. ASSP, June 1983.

[3] - L. B. Almeida, Modelamento Espectral Nao-Estacionario da Fala Vozeada, doutor thesis, IST, Lisbon, Portugal, Feb.1983.

[4] - L. B. Almeida and J. M. Tribolet, "A Spectral Model for Nonstationary Voiced Speech", Proc. ICASSP 1982.

[5] - L. B. Almeida and F. M. Silva, "Harmonic Coding with Variable-Frequency Synthesis", Proc. 1983 Spain Workshop on Signal Processing and its Applications, Sitges, Spain, Sept. 1983.